LinkWinds, A Visual Data Analysis System And Its Application To The Atmospheric Ozone Depletion Problem

ALLAN S. JACOBSON
budj@apex.jpl.nasa.gov

ANDREW L. BERKIN
berkin@krazy.jpl.nasa.gov
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109

The Linked Windows Interactive Data System (LinkWinds) is a prototype visual data exploration system resulting from a NASA/JPL program of research into applying graphical methods for rapidly accessing, displaying and analyzing large multivariate multidisciplinary data sets. Running under UNIX, it is an integrated multi-application executing environment using a data-linking paradigm to dynamically interconnect and control multiple windows containing a variety of displays and manipulators. This paradigm, resulting in a system much like a graphical spreadsheet, is not only a powerful method for organizing large amounts of data for analysis, but leads to a highly intuitive, easy-to learn user interface. It provides great flexibility in rapidly interacting with large masses of complex data to detect trends, correlations and anomalies. The system, containing an expanding suite of non-domain specific applications, provides for the ingestion of a variety of database formats, and hard copy output of all displays. Remote networked workstations running Link Winds may be interconnected, providing a Multi-User Science Environment (MUSE) for collaborative data exploration by a distributed science team. The system is being developed in close collaboration with investigators in a variety of science disciplines using both archived and real-time data, It is currently being used to support the Microwave Limb Sounder (MLS) in orbit aboard UARS. The application of LinkWinds to this data to rapidly detect features such as the ozone hole configuration, and to analyze correlations between chemical constituents of the atmosphere will be described.

1. INTRODUCTION

Recent advances in remote sensing capabilities and computational power are providing unprecedented ability to study our world. These improvements are also producing an ever increasing flood of data which must be gathered, transported, stored, and analyzed to be fully utilized. This paper reports on a system capable Of facilitating the rapid visual analysis of large masses of data, and discusses its application to upper atmospheric science. This system grew out of a NASA project at the Jet Propulsion Laboratory to study the application of computer graphics to the problems of quickly and interactively exploring and analyzing very large amounts of scientific data. The objectives of the program are (1) to develop a software environment which will support the rapid prototyping of visual data analysis applications, while at the same time maintaining the high level of performance necessary for interactively manipulating graphical displays; (2) to develop a user interface that is truly intuitive, allowing quick access to the software for the novice as well as the advanced user; (3) to provide a suite of sample applications which are useful across a variety of scientific disciplines; and (4) to provide tools to support user development of applications for this environment.

2. LINKWINDS

The Linked Windows Interactive Data System, or LinkWinds, is a prototype product of this research effort. In compliance with the research objectives, LinkWinds, an integrated multi-application execution environment w full graphical user interface (GUI), is a visual data analysis and exploration system designed to rapidly and interactively investigate large multivariate and multidisciplinary data sets to detect trends, correlations and anomalies. The system, operating under UNIX, is based on an object-oriented programming model and is implemented in the C language. It draws upon the Silicon Graphics Inc. (SGI) GL library for its GUI and graphics support software and presently runs only on workstations supporting this library. This includes all SGI workstations, and those of other manufacturers who have licensed and support the GL library. Third party software and the advent of OpenGL as a standard graphics library will greatly increase the portability of LinkWinds in the near future.

Data sets and individual tools for display or control of the data are coded as objects, each occupying a window on the LinkWinds screen, and communicating with other objects through a message passing protocol. The objects 01 windows can be linked or unlinked at the discretion of the user. Linking the windows sets up one-way message paths between objects. This data-linking paradigm makes the system perform much like a graphics spreadsheet, and as in a spreadsheet, is a powerful way of organizing the data for analysis while providing a natural and intuitive interface. Data-linking, and its user interface implications, are discussed below.

Messages generated by LinkWinds objects are recorded as program statements in an underlying language called Lynx. The message passing characteristics are the basis for two key LinkWinds functions. The first is the maintenance of an internal journal of all user originated commands executed by the environment. This file can be saved at any time through a menu option. The record can then be replayed at the initiation of subsequent LinkWinds sessions, allowing the user to draw upon a previous layout, of LinkWinds applications and links, or repeat a full analysis session.

The second function based upon the Lynx message passing protocol is the Multi-User Science Environment (MUSE), which provides a method for multiple LinkWinds systems to communicate via networks. Using menu options, users remotely separated can connect to one another, and by also establishing a telephone voice connection, can cooperatively view and manipulate their data. A successful connection requires that each user be executing LinkWinds and that each has access to the data sets being analyzed. This is normally arranged by transporting the data sets prior to the collaborative session. Because only the messages are sent, and not the actual data, a very low bandwidth is required, making for quick and efficient communication. The MUSE capability is also used to give tutorials over the network to new users and to allow users to demonstrate recommendations for application changes or to point out bugs.

Hard copy of the LinkWinds displays are provided by function keys on the keyboard. Placing the cursor in a window and pressing F1 produces an image of a window's contents; pressing F2 saws the complete window and frame; and F3 saves the full screen. The figures shown were obtained in this manner.

3. DATA-LINKING AND THE USER INTERFACE

In addition to the normal GUI functions provided by the windowing environment, dynamic manipulation of graphs and images is facilitated through the data-linking paradigm. Data-linking can be understood in the context of a spreadsheet, where cells containing numbers are linked to other cells. Formulae are associated with each cell, so that when a number changes, all cells linked to the changed cell recalculate their value.s. LinkWinds does the same thing, but in a graphics environment where the rigid grid structure gives way to free form, and a cell can translate, for instance, into sliders or large scale number arrays such as images.

This data-linking paradigm is one of the most distinguishing features of LinkWinds, and evolved from a desire to create a truly easy-t,c)-lcar]l and intuitive user interface. A guiding principle is that users are impatient and want to get started on productive work as quickly as possible. Large manuals only discourage them [M. Rettig, 1991]. Therefore, an interface was needed which can be learned by exploration, and which conforms to expectations as the user works.

Data-linking is affected through two icons. The link icon is a button displaying two interlocking rings, while the unlink icon displays two rings that are separated. Objects on the screen may have a single link button, the full set of link and unlink buttons, or no buttons. The presence of a single link button indicates a data object, while the presence of the pair indicates applications with control functions. A window with no buttons is an application with only display capabilities. To perform a link, the cursor is placed on the appropriate button, and a "rubber band" is dragged out and dropped into the application to be linked. To break the link, the same thing is done using the unlink button. The rubber bands signifying the links may be either continuously displayed or hidden during the session depending on he user's There are two simple rules to follow in applying the inking paradigm.

1. When as a result of menu selections an empty window appears on the screen, pu data in it. This is done by linking a data object into the window.

2. When an object with the pair of link symbols appears, exercise its control function by linking it into any application object.

Our experiences with users has confirmed the intuitiveness of this paradigm. Scientists very quickly become familiar with the manipulations that control LinkWinds, often grabbing the mouse out of our hands during demonstrations to try themselves. New users typically require about 30 minutes of demonstration to understand the system well enough to embark upon productive work. Recently, groups of incoming Caltech graduate students were sent for a LinkWinds tutorial, and they very quickly were performing competently.

While the LinkWinds scheme of linking together objects on the screen is reminiscent of data-flow systems, the similarity is only superficial. LinkWinds is a multi-application execution environment and therefore has significant architectural differences which result in more efficient use of computer memory, a flatter learning curve and optimization for execution, rather than for programming.

4. DATABASE INTERFACE

The current version of LinkWinds interfaces with both archived and real-time data. In the archived data mode, the primary data interface is with the widely accepted 8-bit raster and scientific Hierarchical Data Formats (HDF), created and supported by the National Center for Supercomputing Applications (NCSA) at the University of Illinois, at Urbana-Champaign. The 8-bit raster format allows 3-dimensional data files to be constructed as a sequence of images. LinkWinds also accepts data in a byte format, the Silicon Graphics Inc. RGB image format, and the Common Data Format (CDF), originated and supported by the Goddard Space Flight Center (GSFC). Additional data formats are imported using DataHub [T.H. Handley, Jr. and M.R. Rubin, 1993], a system being developed in close collaboration with the LinkWinds effort, and also supported by the NASA Applied Information Systems Research Program. It is currently used to provide LinkWinds direct access to a wide variety of data formats including the Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) format, Planetary Data System (PDS) format, and NetCDF, a derivative of CDF developed and supported by the National Center for Atmospheric Research (NCAR). The two projects are working together so that DataHub and LinkWinds may be called from inside the other, thus complementing their capabilities. LinkWinds will use DataHub for additional formats and to reduce large data sets, while DataHub will utilize LinkWinds for visual data selection, subsetting and validation.

Data sets to be ingested by LinkWinds are listed in a text file and appear in the top level "Databases" menu. Wild cards may be used in this text file, allowing whole classes of data files to be accessible. New data files created during a session by DataHub or LinkWinds' subsetting tool may be automatically added to the menu. Metadata needed to translate data and axis values to meaningful numbers, such as the number of axes and their names, is obtained from data files in the standard formats. A special metadata file exists for the raw byte data or to provide additional information not contained in the standard formats.

5. APPLICATION TO ATMOSPHERIC DATA

A suite of applications useful across many disciplines has been developed for the LinkWinds environment (see Table 1). Figure 1 shows a typical session to explore a data set collected by the Microwave Limb Sounder (MLS currently in orbit aboard the Upper Atmosphere Research Satellite (UARS) [Waters et al., 1993]. This data is three dimensional, with the concentration of various chemical constituents varying with latitude, longitude, and pressure. The instrument has sensitivity over the range from 100 millibars (mb) to .1 mb. A new data set is collected daily; the data shown here is from day 56 of the mission, which is day 310 of 1991.

The LinkWinds top-level menu appears in the center left of the figure. From this menu the databases, tools and system options are selected. Data objects, with their single link buttons, are just above the menu. In this case, the data displayed are ozone and water vapor. The window entitled Image1 contains a slice of the ozone data at a pressure of 21.54 mbars, as selected by Slider1, which is linked to it. Slider1 is also linked to Image2, which shows the water vapor at the same pressure. Slider1 permits the user to scan the full data set from the maximum to minimum altitudes. The user can also switch to any of the three orthogonal axes and similarly scan them. The amount of each constituent is given by color, as indicated on the color bars at the bottom of the Images, with red denoting high concentration and blue low. Grey indicates missing data, where MLS could not measure due to the orbital position, orientation and sensing range of UARS.

The southern hemisphere ozone hole is clearly seen at the lower left of Image1, and a high value of the water vapor is seen in the corresponding location of Image2. This anticorrelation is readily visible in Scatter1, where the points shown come from the region defined by a bounding box control embedded in Image and linked to Scatter1. This box, visible in the lower left of Image1, can be resized or moved as desired. For every point inside this box, Scatter1 plots the ozone vertically versus water vapor horizontally. In the text box of Scatter1, statistical quantities associated with the scattered data are given. Of particular note is the closeness of the correlation coefficient to - 1, indicating high linear anticorrelation. Other information such as third and fourth moments, linear best fit data of the scatter, the number of points scattered, and the chi-squared value may be selected from the Scatter menu. The full range of statistical information may be saved to a file via a button.

Two other applications display data along one dimension. In the lower center of Figure 1, LinePlot1 displays ozone concentration as a function of pressure. The location in latitude and longtitude of this plot is controlled by Image1, which has been linked into LinePlot1. The green plot corresponds to the frozen crosshair at the center of Image1, while the white plot corresponds to the red crosshair and is instantaneously updated as the crosshair is moved. The much lower white values reflect the paucity of ozone inside the hole. The Profile application to the left of LinePlot1 shows ozone concentration along the cyan line in the slice of Image1 This line may be interactively updated. The current Profile line starts in the ozone hole, goes through an ozone rich region, and ends back in the hole. This is reflected both in the colors of Image1 as well as the height of the curve of Profile1, providing a different way of visualizing the data.

Because this data Set is global, Figure 2 shows the ozone data displayed in Globe1 as both a color and height field rendered on a sphere. The ozone hole, and adjacent regions of higher ozone, are clearly seen. The water vapor data could have been used for the height field, clearly demonstrating any correlations. Slider1 also sets the pressure slice of this display, and the height scale is determined by the rotary dial on the left side of the window. Pan-Zoom and 2-Axis Rotator controls linked to Globe1 allow it to be interactively positioned as desired.

The Combine tool allows mathematical manipulations to be performed on slices of data. A calculator called from Combine via a button contains the standard mathematical functions, and allows the input of constants as well as slices. The user can combine up to three slices of data, from the same set or mixed sets, in any mathematical expression desired. In Figure ?, these slices have been selected by linking in LinePlot, with the red, green, and blue sliders setting the levels to be 21.54, 14.68, and 10 mb. The calculator has summed these three slices, and the resulting slice and histogram of the data distribution are displayed in Combine1

Among the tools in Table 1, two which deserve special comment facilitate the creation of animations for immediate display on the screen or subsequent recording on video or film. One animator is frame based, with the user selecting starting and ending Control values, and the number of frames desired in the animation. The other animator is time based. The user sets any number of control positions, each with associated key times graphically selected by moving the hands on a clock. The Animator then interpolates between these set positions to easily make animations with great flexibility in the control, The desired frame rate is selected from the animator menu. These include those for film, video, and a screen display mode which makes the animation in real time.

6. FUTURE PLANS

Several developments are planned for the future to significantly improve the usefulness of LinkWinds. As in the past, we will continue to develop applications in collaboration with science users seeking to solve real problems. Where relevant, these applications will make use of modern rendering techniques which can successfully be applied in an interactive environment.

A major impedance to the use of any visualization tool is the difficulty users have in inputting their data. Developments of LinkWinds and DataHub will continue to make this process as seamless and automatic as possible. A related issue is the types of data addressed by LinkWinds. Available tools for visual data analysis are generally confined to relatively well-behaved and rectangularly gridded data sets. LinkWinds' first venture from the common mold was into the realm of real-time data, creating a capability to ingest such data and building interactive applications for monitoring and analyzing it. There are other major neglected categories of data that are quite common in scientific research, and badly in need of tools to support their exploration and analysis. Several problem areas to be addressed in the future are (1) data sets in which there are significant sources of error, either statistical and/or systematic; (2) data sets which are ungridded samples, either sparse or numerous, from which the user desires to construct gridded data sets over extended regions; and (3) disparate sized data sets from a variety of instruments which must be warped and/or co-registered for overlay or comparison.

In the future, we intend to pursue the development of a users' applications generator for LinkWinds. Currently, the layout of the objects, or widgets, in all of the windows is determined by a text file. The user can re-configure these windows either by editing this file or interactively from a menu-selectable "redesign" mode. We intend to expand this tool kit approach to further allow users to throw widgets away, or add new widgets from a provided catalog. In conjunction with the widget moves, LinkWinds will generate a C code source module to make the widgets work. This code will be suitable for use as a template for the development of a full application. As experience is gained with this approach, and a planned conversion of the code to C++ is accomplished, we anticipate that the rendering and display processes will also lend themselves to a limited catalog of processes selectable by the user.

Aware that the only way to develop useful tools is in conjunction with research on meaningful problems, our policy has been to encourage users and potential users to contact us concerning LinkWinds' changes and needs. We have responded and will continue to do so to the limit of our resources. LinkWinds is currently in use at more than ten institutions, being applied to problems in both remote-sensed and field geology, atmospheric physics and chemistry, meteorology, oceanography, chemical spectroscopy, space plasmas, genetics and cellular biology. As it evolves we expect to interact with a wider distribution of scientists engaged in research spanning additional scientific disciplines.

Acknowledgments. In addition to the authors, the current LinkWinds development team members include Martin Orton and Mark Rubin at JPL, and Philip Mercurio and James McLeod at the San Diego Supercomputer Center. Alumni who made important contributions are Brian Beckman, Leo Blume, Bonnie Boyd and Mitch Wade. Lee Elson provided useful discussions on upper atmospheric science. ALB acknowledges support from the National Research Council. We extend our thanks to Joe Bredekamp and his staff at NASA, whose sponsorship makes this work possible. The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

Handley, T.H. Jr., and M.R. Rubin, DataHub: Intelligent distributed science data management to support interactive exploratory analysis, in Computing in Aerospace 9 (in press), AIAA, Washington, D.C., 1993.

Jacobson, A. S., and A. L. Berkin, LinkWinds, a system for interactive scientific data analysis and visualization, in High Performance Computing 1993, edited by A. Tentner, pp. 214-219, The Society for Computer Simulation, San Diego, CA, 1993.

Rettig, M., Nobody reads documentation, CACM 34, 19-24, 1991.

Waters, J. W., L. Froidevaux, W.G. Read, G.L. Manney, L.S. Elson, D.A. Flower, R.F. Jarnot, and R.S. Harwood, Stratospheric ClO and ozone from the Microwave Limb Sounder on the Upper Atmosphere Research Satellite, Nature, 362, 597-602, 1993.

	Control
Slider	Controls which slice of data is displayed along any of three orthogonal axes
3-Axis Slider	Three sliders control the slice of data displayed along the three orthogonal axes
3-Axis Rotator	Three sliders, one for each axis, rotate three dimensional applications
2-Axis Rotator	One slider moving in a two dimensional area rotates three dimensional applications
Pan-Zoom Slider	Controls three dimensional applications by changing the viewpoint
Combine Slider	Six sliders determine three slices of data and their three absolute offsets
Animator	Provides time based animations of any application with an arbitrary number of key frames, with a variety of frame rates
Frame Animator	Provides frame based animations of any application, with only the start and stop control values being set
Color Tool	Interactive data set palette manipulation allowing color editing, data ranges being ramped in color, or substitution of pre-defined palettes
	Display/Control
Combine	Up to three slices of data are combined using standard mathematical functions entered in an embedded calculator
Compare	The functional behavior of each point in a data set is compared with a reference point using a variety of mathematical functions
Data Subset	Allows the user to interactively save portions of the displayed data into HDF
LinePlot	Plots the values along a straight line going completely through a data set parallel to any axis, and also functions as a slider to select three slices
Histogram	Displays the distribution of values in the 256 data channels for up to three slices, and provides filtering and color stretching
Image	Displays a single slice of data or a composite RGB image of three slices with embedded crosshair, bounding box, and line controls
	Display
Plane	Renders an image in perspective relief, with an optional accompanying height field, of either a single slice or RGB three slice composite
Globe	Renders an image on a globe, with an optional accompanying height field, of either a single slice or RGB three slice composite
Orthoview	Displays in a three dimensional rendering all the points in a data set between two values
Profile	Displays the data values along a line drawn on the Image, Combine, or Compare tools
2D Scatter Plot	At every location in a slice, plots the values of one data set against the other to show the correlation
3D Scatter Plot	At every location in a slice, plots the values of three data sets in three dimensions to show their correlation
TrackPixel	Gives numerical information about the data in Image, Combine, or Compare both at a point and averaged over a bounding box
	Real-time
StreamPlot	A strip-chart recorder displaying data as either color or line plots as a function of channel number vertically and time horizontally
StreamLine	Plots data value versus channel number, updating in real time with options of saving a spectrum and averaging in time
StreamPlane	Functions much like StreamPlot, but with the data rendered in relief using the data values for height and color
StreamClock	Provides timing information for the current data, giving date, day of year, time of day, and internal spacecraft time
ChannelSlider	Controls the range of channels to be viewed, allowing concentration on features of interest

Table 1. Current Suite of LinkWinds Applications

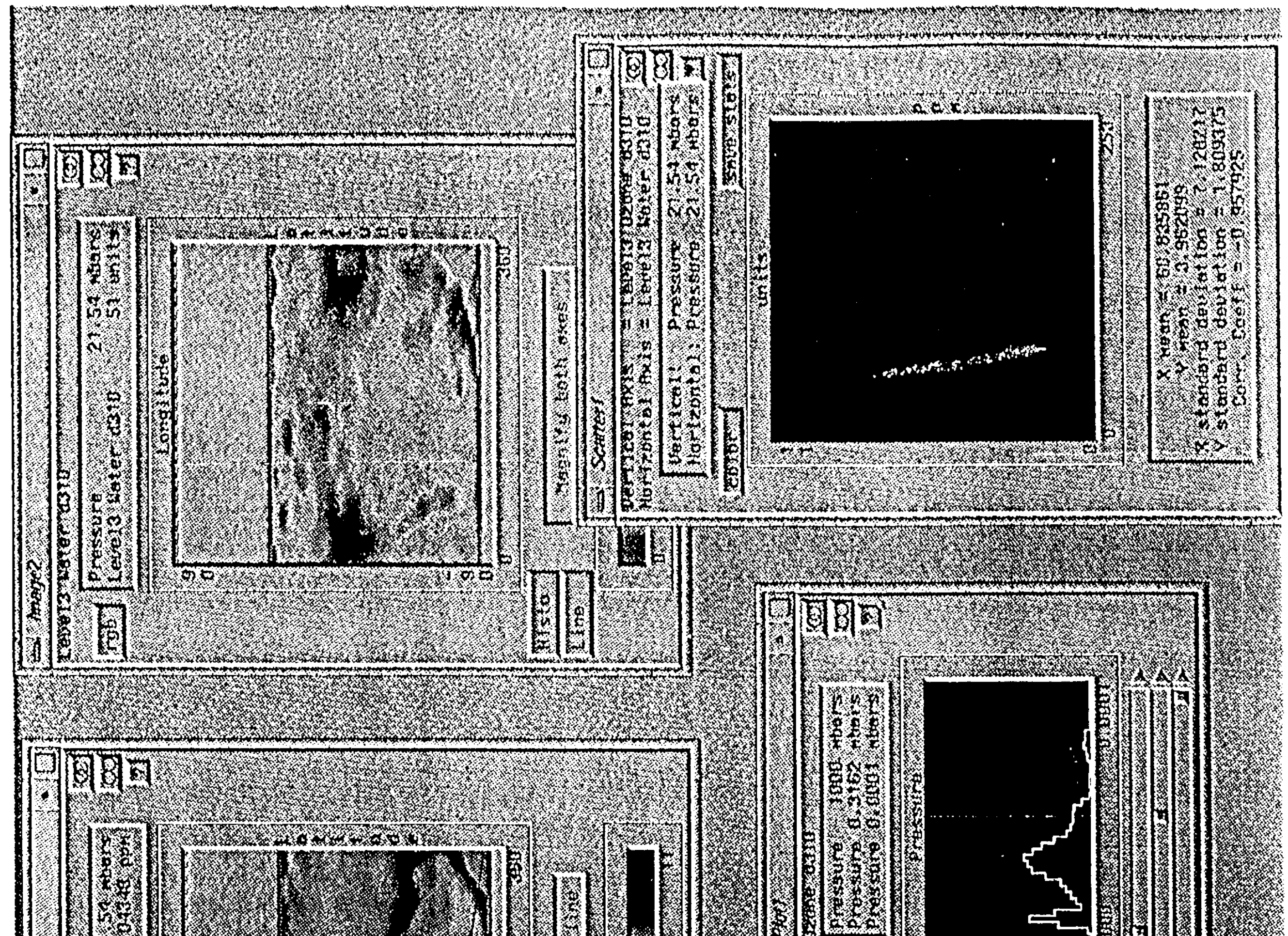

Figure 1. LinkWinds session to explore upper *atmospheric ozone and water vapor measured by the Microwave Limb Sounder aboard UARS*

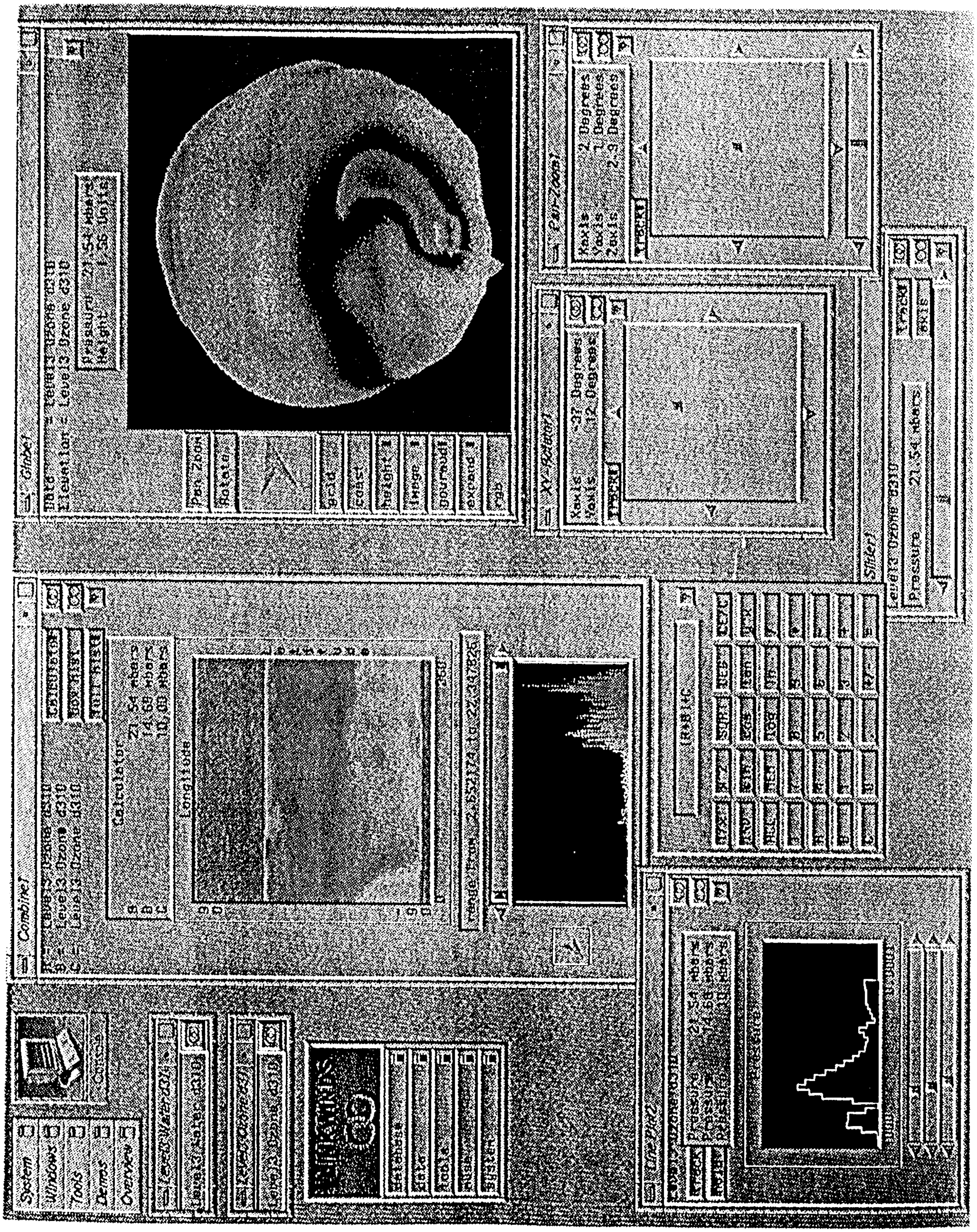

Figure 2. Examining other features of upper atmospheric ozone measured by the Microwave Limb Sounder aboard UARS